The First Large Scale Audio-Visual Dataset of Cognitive Load and Affect For Remote Work

Pritam Sarkar and Ali Etemad

Dept. of Electrical and Computing Engineering & Ingenuity Labs Research Institute
Queen's University, Kingston, Canada

Introduction

Motivation. Due to recent COVID-19 pandemic, *remote work* is the new reality of work for millions across the world. While this new paradigm has a number of advantages such as enabling social distancing and flexible hours, it brings about a number of challenges that were less common in in-person work environments. For instance, studies have shown that *remote work settings could contribute to increased cognitive load and fatigues in individuals* due to the reasons including but not limited to:

- back-to-back work-related meetings with minimal physical mobility in-between,
- the inability to effectively perceive and transmit non-verbal expressive cues,
- the need to apply intense focus on the screen with minimal variation.

In order to better understand mental health and manage the impact of remote work meetings on individuals, it is necessary to design and develop tools capable of quantifying factors such as cognitive load and affect in relevant settings.

Problem Statement. A key ingredient for developing such systems is the availability of related *datasets* along with *ground-truth information*, which could be used by AI researchers and psychologists to acquire better understanding of cognitive load, affective states, and broadly human behaviors in this new paradigm of work. In best of our knowledge, there are currently no available datasets directed toward understanding users in remote work settings. *Moreover, there are currently no audio-visual datasets that target cognitive load along side affective states.*

Our Contribution. In this work, we fill an existing timely gap by introducing **AVCAffe**, the first **A**udio-**V**isual dataset consisting of **C**ognitive load and **Affe**ct attributes. AVCAffe would be of interests to both the AI researchers and psychologists to model and test the hypotheses related to the impact of varying amount of cognitive load and affective states in remote work.

AVCAffe

Objective. We aim to design a study protocol that closely resemble remote work and meetings. To facilitate this concept, we devise a series of tasks with varied levels of difficulty, eliciting cognitive load and affect in a mimicked remote work setup.

Study Design. Our study design requires 2 participants to collaborate and communicate over a video conferencing platform to successfully complete a series of tasks including:

- Open discussion: conversation on non-personal topics like movies, books etc.
- Lighten the mood: sharing some interesting or humorous incidents, jokes etc.
- Diapix: finding differences between 2 highly similar pictures.
- Montclair map: locating the missing landmarks of 2 nearly identical maps.
- Lost at sea: focus on decision making skill in a hypothetical scenario.
- Reading comprehension: read a passage and answer some questions.
- Multi-task: email writing with interruptions unrelated to the email.

We carefully choose these tasks to target varying levels of cognitive and affective states at different stages of the experiment.

Data Acquisition. We capture the audio-visual recordings of the participants through out the session, along with self-reported ground truths at the end of each task. In particular, cognitive load and affect scores are collected using the NASA Task Load Index and Self Assessment Manikin. We present representative frames and ground truths reported by two of the participants in Fig. 1.

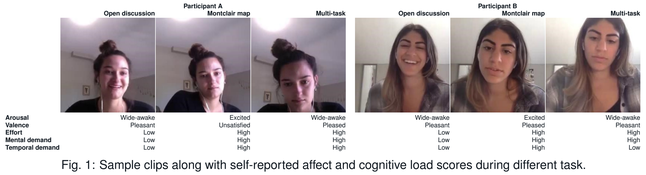

Fig. 1: Sample clips along with self-reported affect and cognitive load scores during different task.

Dataset Statistics. We summarize the key statistics of the AVCAffe below.

# Subjects. 106	**Duration.** 108 hrs.	**# Clips.** 58,112
Gender. Male: 52, Female: 53, Non-Binary: 1.	**Age:** 18 to 20 : 8; 21 to 30 : 75; 31 to 40 : 17; 41 to 50 : 2; 51 to 60 : 4.	
Countries of origin: Bangladesh(1), Brazil(2), Canada(67), China(3), Ecuador(1), Egypt(1), Germany(1), Hong Kong(1), India(11), Iran(4), Ireland(1), Jordan(1), Mexico(4), Nigeria(2), Pakistan(2), Sweden(1), USA(2), Vietnam(1)		
Ground truths. Arousal, Valence, Mental Demand, Temporal Demand, Effort, Physical Demand, Performance, and Frustration		

Analysis

We perform an in-depth analysis of self-reported ground truths to validate the success of our study design, indicating that different tasks are able to induce varying degrees of cognitive and affective states amongst the participants throughout the session.

Cognitive Load. While analyzing cognitive states, we present the density plots of each cognitive load attribute such as, mental demand, effort, and temporal demand in Fig. 2. We find distinct shifts in cognitive load over time, specifically for *mental demand*, *effort*, and *temporal demand*. For example. participants' *mental demand* is fairly low during the *open discussion* and *lighten the mood* but report higher mental demand during *multi-task* and *reading comprehension (active)*.

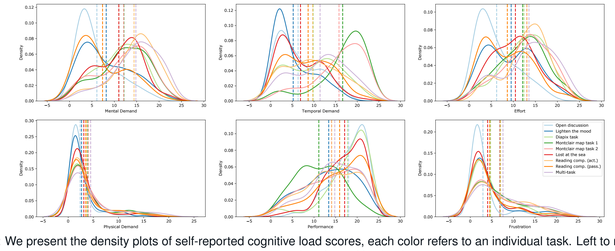

Fig. 2: We present the density plots of self-reported cognitive load scores, each color refers to an individual task. Left to right in an increasing order of cognitive load.

Affect. To analyse the self-reported affect scores, we project arousal and valence responses onto a 3D valence-arousal space, presented in Fig. 3. We notice considerable shifts in self-reported responses across different stages of the experiment. For example, during *Open discussion* and *Lighten the mood*, majority of the participants experience 'pleasant' and 'excited', whereas, during *multi-task*, majority of the participants report 'unsatisfied' and 'wide-awake'.

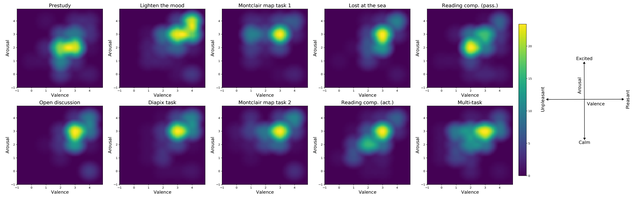

Fig. 3: The affective scores projected in a 3-d plane, where the colors denote population density, yellow being the most dense.

Cognitive Load vs. Affect. In Fig. 4, we project the self-reported ground-truths onto a 3D space to study the relationships between different affect and cognitive load attributes. We observe a strong positive correlation between *effort* and *mental demand*. Moreover, some degree of positive correlation is noticed between *temporal demand* and *effort*, as well as, between *temporal demand* and *mental demand*. Interestingly, we do not observe strong correlations between cognitive load and affect attributes, which further proves that our dataset has been able to successfully capture unique information beyond the arousal and valence classes.

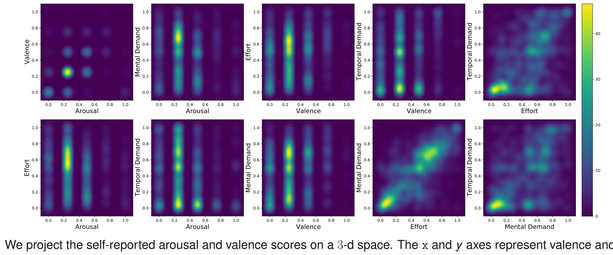

Fig. 4: We project the self-reported arousal and valence scores on a 3-d space. The x and *y* axes represent valence and arousal respectively, and color denotes population density with yellow being the most dense.

Classification of Cog. Load and Affect

Next, we utilize the collected data to train deep learning models towards identifying participant's cognitive load and affect at remote work. A simple illustration of the deep learning model is presented in Fig. 5. Following a standard practice, we use the weighted F1-measures as the evaluation metric because of it's robustness towards imbalanced class distribution. We notice that multi-modal networks outperform the uni-modal variants on all the attributes.

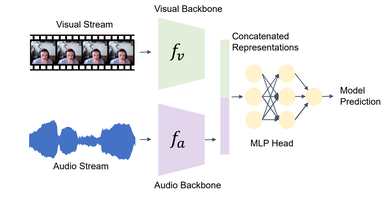

Fig. 5: A simple illustration of deep learning approaches in classifying cognitive load and affect.

Audio	Visual	Mental D.	Effort	Temporal D.	Arousal	Valence
Random classifier		45.6	43.6	33.9	26.2	30.3
✓	✗	61.2	62.1	56.7	36.1	39.1
✗	✓	61.0	67.7	59.4	34.0	39.5
✓	✓	65.6	69.4	66.7	40.5	41.7

Tab. 1: Highlights of the models' performance in classifying participant's cognitive load and affect.

Broader Impacts

- AVCAffe enables the researchers to study the impact of remote work in cognitive load and affect, and broadly mental health and human behavior.
- Moreover, it would enable engineers in developing intelligent and better self-management tools with the ability to cater and deliver personalized solutions.
- In addition to understanding cognitive load and affect at remote work, AVCAffe can be further used to conduct research in several other areas like audio-visual speech recognition, lip reading from visual input, long video summarization, among others.
- The authors do not foresee any major negative societal impact, however, all ethical considerations that apply to other audio-visual affective datasets may equally apply to AVCAffe.

Summary

We present a novel audio-visual database of cognitive load and affect collected in a setup resembling 'remote work'. To the best of our knowledge AVCAffe is the **first audio-visual dataset comprised of *cognitive load*** annotations. Moreover, AVCAffe is the **largest affective computing dataset** in English language. We believe AVCAffe would be a valuable and challenging benchmark for the deep learning and affective computing research communities to accurately model cognitive load and affect, especially considering the **timely context** of remote work.

Additional Information

Manuscript: *Sarkar, P., Posen, A., & Etemad, A. (2022). AVCAffe: A Large Scale Audio-Visual Dataset of Cognitive Load and Affect for Remote Work. arXiv preprint arXiv:2205.06887.*
Paper link: https://arxiv.org/pdf/2205.06887.pdf
This dataset along with supporting codes are made freely available on the project website: https://pritamsarkar.com/AVCAffe/.
You may direct any questions or additional queries at: pritam.sarkar@queensu.ca.